

Smart Computing Card SC7 FP150



Great Computing Power

- Peak performance 154 TOPS INT8
- Peak performance 76.8 TFLOPS FP16/BF16
- Support INT8/FP16/BF16/FP32 mixed precision calculations

Excellent Encoding and Decoding Capabilities

- 192 channels 25fps HD video hardware decoding
- 72 channels 25fps HD video hardware encoding
- Up to 8K resolution of video and image decoding

Wide Adaptation

- Adapt to various servers
- Adapt to various operating systems and mainstream algorithm frameworks

Smart Computing Card SC7 FP150 is a computing power card for cloud AI applications. It can be used for large model inference, and can efficiently adapt to mainstream AI algorithms in the market to realize audio and video auditing, audio and video transcoding, structuring, behavior analysis and other AI applications. These features empower various fields such as Internet, smart cities, smart transportation, smart energy, smart finance, smart telecommunications and smart industry.

Application Scenarios



AIGC



Smart Transportation



Smart Power



Smart Oil and Gas



Smart Cities



Smart Business Hall



Smart Manufacturing



Smart Warehousing

Product Specifications

Product Number		SC7 FP150
CPU		48-core A53 338560 DMIPS 2.3GHz
Deep learning computing power	INT8	154 TOPS peak performance
	FP16/BF16	76.8 TFLOPS peak performance
	FP32	9.6 TFLOPS peak performance
Video/image coding and decoding performance	Video decoding capacity	H.264 & H.265: 1080P @4800fps
	Video decoding resolution	8K / 4K / 1080P / 720P / D1 / CIF
	Video coding capacity	H.264 & H.265: 1080P @1800fps
	Video coding resolution	4K / 1080P / 720P / D1 / CIF
	Image coding/decoding capacity	JPEG: 2400 images / second @1080P (interleave off)
	Image coding/decoding maximum resolution	32768 * 32768
Memory	Standard configuration	96GB LPDDR4x 768bit 328GB/s
PCIe	Data channel	PCIe Gen4 X16
Power consumption	Maximum power consumption	150W
Working environment	Temperature range	0°C ~ 45°C
	Method of heat dissipation	Passive
Structure size	Length * height * width	Full height, full length, and single width
Deep learning frameworks		TensorFlow / PyTorch / Paddle / Caffe / ONNX / MXNet / DarkNet
Operating system support		Ubuntu / CentOS / Debian / UOS /Kylin
Main control CPU support		Compatible with Intel、AMD、Feiteng、Haiguang、Shenwei、Loongson, etc

Application-optimized software SDK



- Support various deep learning frameworks: TensorFlow, PyTorch, ONNX, PaddlePaddle, Caffe, etc
- Built-in 150 operator implementations and user-defined operators
- Support dynamic tensor size
- Support control flow networking
- Support user quantitative model import
- Support automatic quantitative strategy search
- Supports immediate data verification during compilation



- Efficiently adapt to various deep learning algorithms Support large model reasoning, such as ChatGLM, LLaMA, Falcon
- Support AI drawing, SD V1.5, SDXL
- Support various image classification and detection networks, such as Resnet50, Yolo series
- Support various natural language processing networks, such as BERT series
- Support various voice networks, such as Tacotron2



- Provides powerful development libraries (BMCV、FFMPEG coding and decoding interface、OPENCV coding and decoding interface、OPENCV extension)
- Support Docker containerization, Kubernetes extension management

Framework

PyTorch

TensorFlow

Paddle Paddle

Caffe

TFLite

Frontend

OnnxConverter

CaffeConverter

TFLiteConverter

Compiler

TPU-MLIR

Runtime

BMRuntime Engine

BMCV

OpenCV

FFMpeg

BMLib

Drivers

Unified cloud-edge-end tool chain